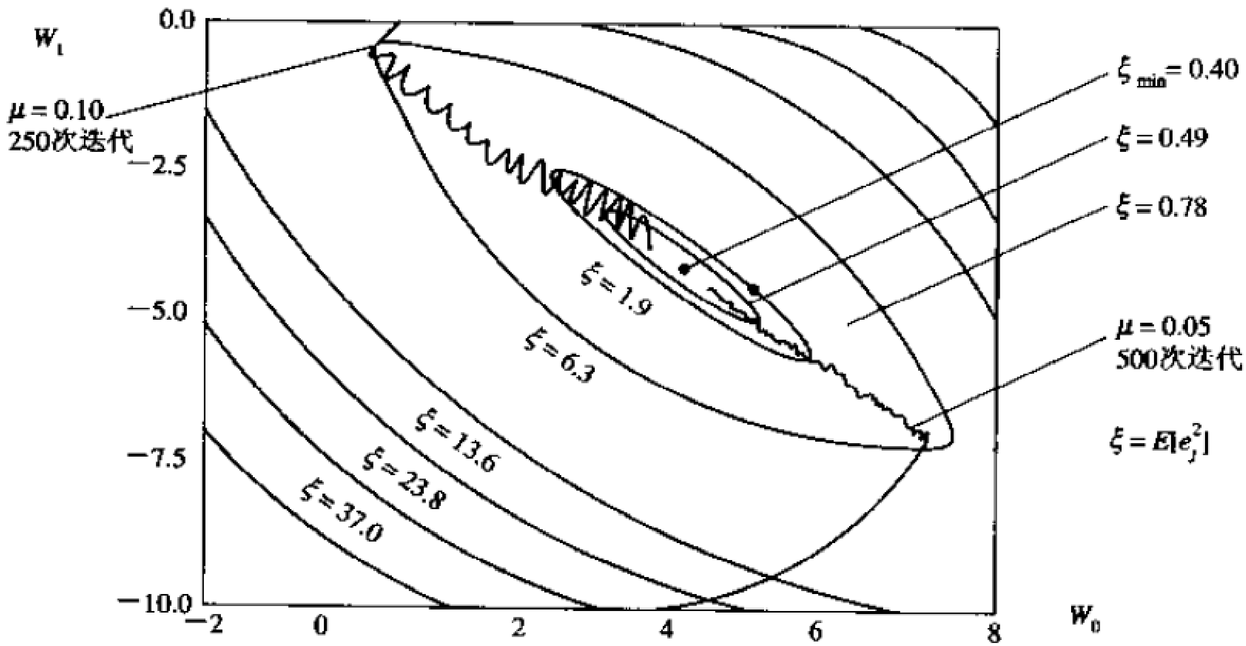
- 权值向量为二维的情况

✓ 大步长($\mu=0.1$)：收敛速度快，随机抖动性大，收敛后精度低；

✓ 小步长($\mu=0.05$)：较为平稳，精度高，收敛缓慢



3. LMS算法的权值期望 $E[\mathbf{w}(n)]$ 性能

- 将LMS算法的迭代公式展开，可得

$$
\begin{aligned}
\mathbf{w}(n+1) &= \mathbf{w}(n) + \mu[\mathbf{u}(n)e^*(n)] \\
&= \mathbf{w}(n) + \mu\{\mathbf{u}(n)[d^*(n) - \mathbf{u}^H(n)\mathbf{w}(n)]\}
\end{aligned}
$$

- 上式两边取期望，可得

$$
\begin{aligned}
E[\mathbf{w}(n+1)] = E[\mathbf{w}(n)] &+ \mu E[\mathbf{u}(n)d^*(n)] \\
&- \mu E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)]
\end{aligned}
$$

- 将LMS算法的迭代公式展开，可得

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu[\mathbf{u}(n)e^*(n)]$$
$$=\mathbf{w}(n)+\mu\{\mathbf{u}(n)[d^*(n)-\mathbf{u}^H(n)\mathbf{w}(n)]\}$$

- 上式两边取期望，可得

$$E[\mathbf{w}(n+1)]=E[\mathbf{w}(n)]+\mu E[\mathbf{u}(n)d^*(n)]$$
$$-\mu E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)]$$

- 在步长取值很小的情况下，近似有

$$E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)] = E[\mathbf{u}(n)\mathbf{u}^H(n)]E[\mathbf{w}(n)]$$

- 则有 $E[\mathbf{w}(n+1)] = E[\mathbf{w}(n)] + \mu E[\mathbf{u}(n)d^*(n)]$

$$-\mu E[\mathbf{u}(n)\mathbf{u}^H(n)\mathbf{w}(n)]$$

上式可简化成

$$\begin{aligned} E[\mathbf{w}(n+1)] &= E[\mathbf{w}(n)] - \mu\mathbf{R}E[\mathbf{w}(n)] + \mu\mathbf{p} \\ &= E[\mathbf{w}(n)] - \mu\mathbf{R}E[\mathbf{w}(n)] + \mu\mathbf{R}\mathbf{w}_o \\ &= E[\mathbf{w}(n)] - \mu\mathbf{R}\left\{E[\mathbf{w}(n)] - \mathbf{w}_o\right\} \end{aligned}$$

$$E[\mathbf{w}(n+1)] = E[\mathbf{w}(n)] - \mu\mathbf{R}\left\{E[\mathbf{w}(n)] - \mathbf{w}_o\right\}$$

令$\mathbf{v}(n) = E[\mathbf{w}(n)] - \mathbf{w}_o$,上式两边同时减去 $\mathbf{w}_o$得

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \mu\mathbf{R}\mathbf{v}(n) = \mathbf{Q}[\mathbf{I} - \mu\mathbf{\Lambda}]\mathbf{Q}^H\mathbf{v}(n)$$

两边同乘$\mathbf{Q}^H$，上式可化为

$$\mathbf{v}'(n+1) = [\mathbf{I} - \mu\mathbf{\Lambda}]\mathbf{v}'(n)$$

由式 $\mathbf{v}'(n+1)=[\mathbf{I}-\mu\mathbf{\Lambda}]\mathbf{v}'(n)$ 递推，可得

$$\mathbf{v}'(1)=[\mathbf{I}-\mu\mathbf{\Lambda}]^1\mathbf{v}'(0)$$

$$\mathbf{v}'(2)=[\mathbf{I}-\mu\mathbf{\Lambda}]\mathbf{v}'(1)=[\mathbf{I}-\mu\mathbf{\Lambda}]^2\mathbf{v}'(0)$$

$$\vdots$$

$$\mathbf{v}'(n)=[\mathbf{I}-\mu\mathbf{\Lambda}]^n\mathbf{v}'(0)$$

由上式可得如下结论：

- ✓ 每次迭代，权值误差均值以 $[\mathbf{I}-\mu\mathbf{\Lambda}]$ 速度衰减。

权值期望误差指数衰减的时间常数

- ✓ 每次迭代，权值误差均值以 $[\mathbf{I}-\mu\mathbf{\Lambda}]$ 速度衰减。

如果权向量有M个分量，相关矩阵就有M个特征值，则权值的衰减具有：

- ✓ M个特征衰减模式，其

  第m个权值期望以$(1-\mu\lambda_m)$速度衰减。

定义$\gamma_m = 1 - \mu\lambda_m$。如果将$\gamma_m$看成随迭代次数$n$作**指数衰减**，则有

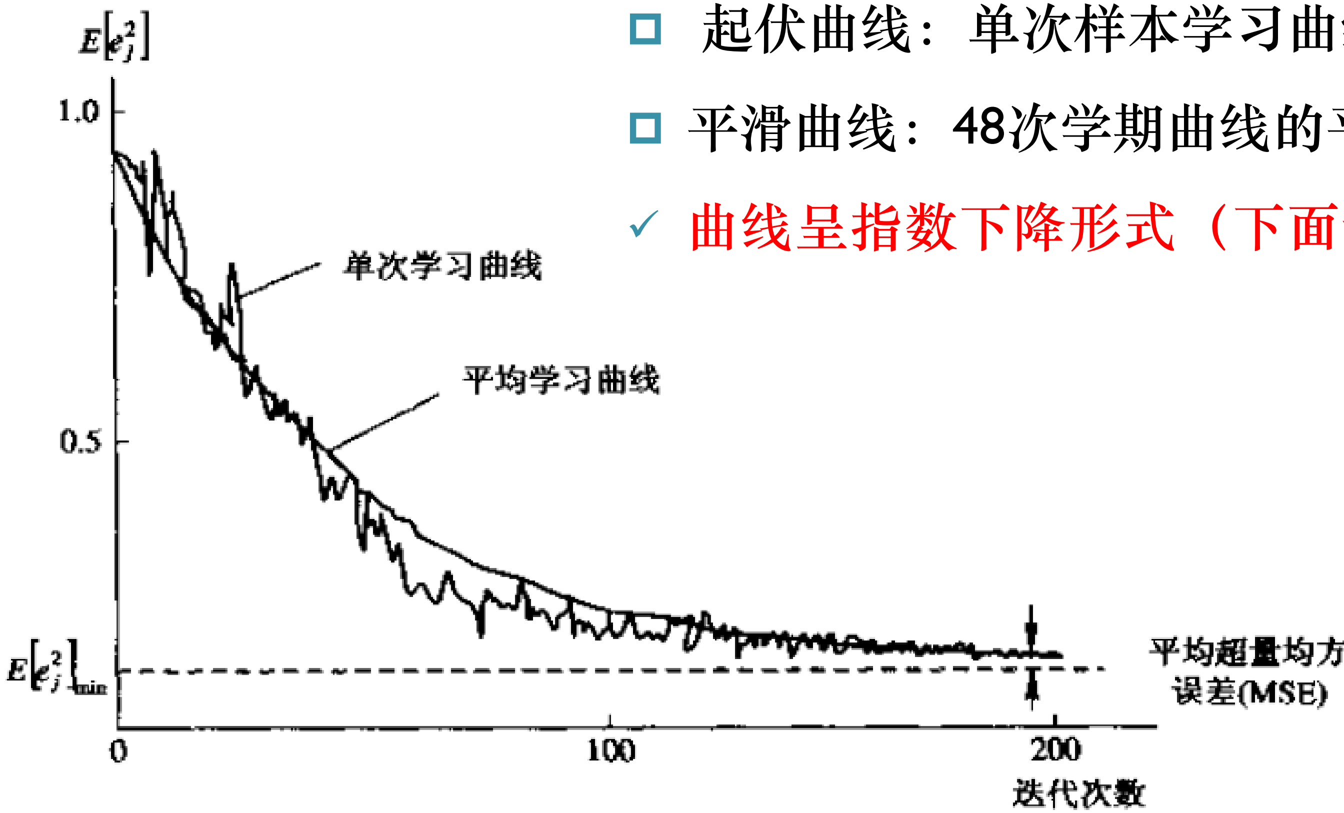
$$\gamma_m = e^{-1/\tau_m} = 1 - \frac{1}{\tau_m} + \frac{1}{2!\tau_m^2} + \cdots$$

对于比较缓慢而平稳的自适应过程，τ_m一般较大，因此可以略去等式右边的高次项，即

$$\gamma_m = 1 - \mu\lambda_m \approx 1 - \frac{1}{\tau_m} \quad \Rightarrow \quad \tau_m = \frac{1}{\mu\lambda_m}$$

4. LMS算法的均方误差 $E[|e(n)|^2]$ 性能

➢ 自适应滤波器的 $\mathbf{w}(n)$需要经过一个迭代过程才能达到$\mathbf{w}_o$，即$E[e^2(n)]$趋于$E[e^2(n)]_{\min}$ 需要一个过程。



- 起伏曲线：单次样本学习曲线
- 平滑曲线：48次学期曲线的平均
- ✓ 曲线呈指数下降形式（下面证明）

$$E[|e(n)|^2] = E\left\{\left[d(n) - \mathbf{w}^H(n)\mathbf{u}(n)\right]^2\right\}$$

$$= E\left[|d(n)|^2\right] - \mathbf{p}^H E[\mathbf{w}(n)] - E[\mathbf{w}(n)]^H \mathbf{p} + E[\mathbf{w}(n)]^H \mathbf{R} E[\mathbf{w}(n)]$$

将维纳解 $\mathbf{w} = \mathbf{w}_o = R^{-1}\mathbf{p}$ 代入均方误差表达式，得

$$E\left[|e(n)|^2\right]_{\min} = E\left[|d(n)|^2\right] - \mathbf{p}^H \mathbf{w}_o - \mathbf{w}_o^H \mathbf{p} + \mathbf{w}_o^H \mathbf{R} \mathbf{w}_o$$

$$= E\left[|d(n)|^2\right] - \mathbf{p}^H \mathbf{w}_o \qquad (\mathbf{p} = \mathbf{R}\mathbf{w}_o)$$

$$\begin{cases} E[|e(n)|^2] = E\left[|d(n)|^2\right] - \mathbf{p}^H E[\mathbf{w}(n)] \\ \qquad\qquad - E[\mathbf{w}(n)]^H \mathbf{p} + E[\mathbf{w}(n)]^H \mathbf{R} E[\mathbf{w}(n)] \\ E[|e(n)|^2]_{\min} = E[d^2(n)] - \mathbf{p}^H \mathbf{w}_o \end{cases}$$

由上述两式，可得

$$E[|e(n)|^2] = E[|e(n)|^2]_{\min} + \mathbf{p}^H \mathbf{w}_o - \mathbf{p}^H E[\mathbf{w}(n)] \\ \qquad - E[\mathbf{w}(n)]^H \mathbf{p} + E[\mathbf{w}(n)]^H \mathbf{R} E[\mathbf{w}(n)]$$

令$\mathbf{v}(n)=\mathbf{w}(n)-\mathbf{w}_o$，进一步转换，可得

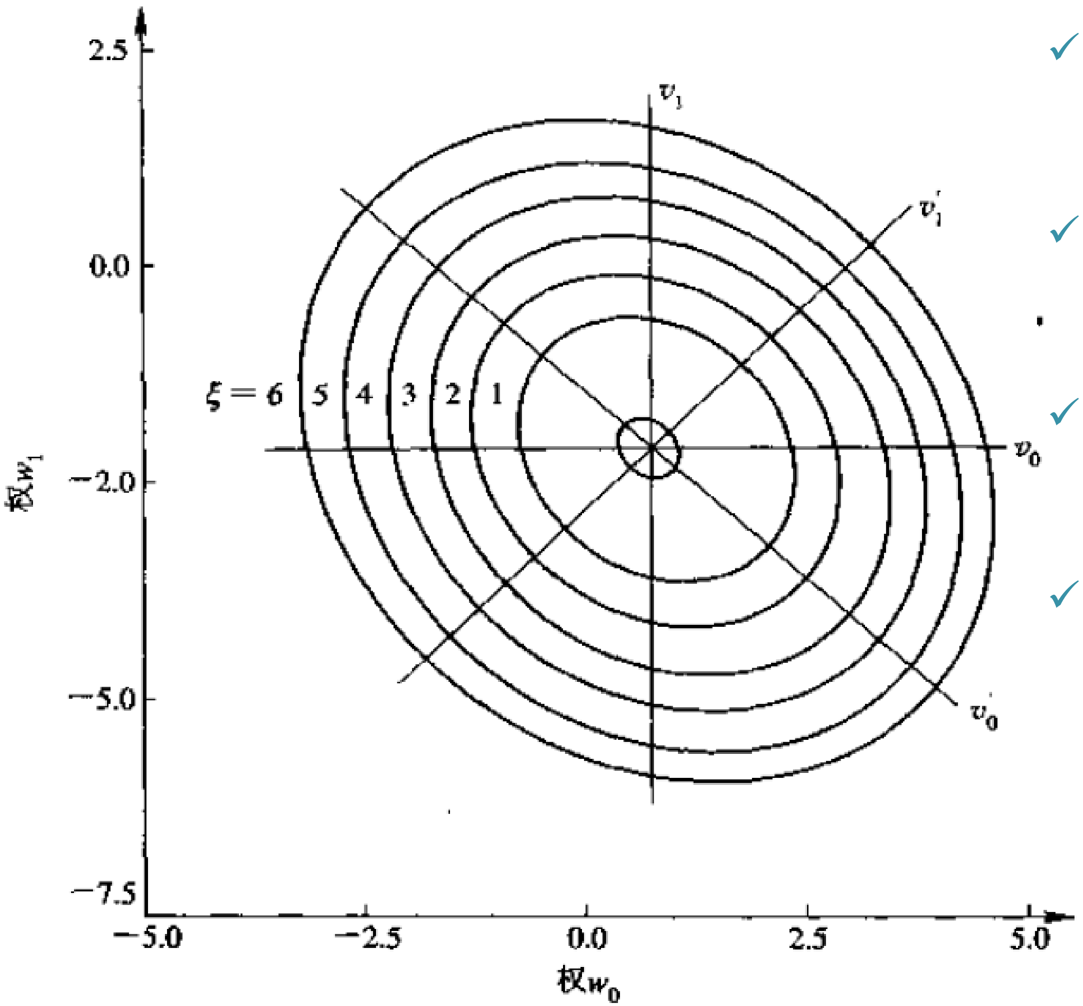
$$
\begin{aligned}
E\left[\left|e(n)\right|^2\right] &= E[|e(n)|^2]_{\min}+\mathbf{p}^H\mathbf{w}_o-\mathbf{p}^H E[\mathbf{w}(n)] \\
&\quad -E[\mathbf{w}(n)]^H\mathbf{p}+E[\mathbf{w}(n)]^H\mathbf{R}E[\mathbf{w}(n)] \\
&= E[|e(n)|^2]_{\min}+\mathbf{w}_o^H\mathbf{R}\mathbf{w}_o-\mathbf{w}_o^H\mathbf{R}E[\mathbf{w}(n)] \\
&\quad -E[\mathbf{w}(n)]^H\mathbf{R}\mathbf{w}_o+E[\mathbf{w}(n)]^H\mathbf{R}E[\mathbf{w}(n)] \\
&= E[|e(n)|^2]_{\min}+\left\{E[\mathbf{w}(n)]-\mathbf{w}_o\right\}^H\mathbf{R}\left\{E[\mathbf{w}(n)]-\mathbf{w}_o\right\} \\
&= E[|e(n)|^2]_{\min}+\mathbf{v}^H(n)\mathbf{R}\mathbf{v}(n)
\end{aligned}
$$

根据特征值分解，令 $\mathbf{R}=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$，则有

$$
\begin{aligned}
E\left[\left|e(n)\right|^2\right] &= E[|e(n)|^2]_{\min} + \mathbf{v}^H(n)\mathbf{R}\mathbf{v}(n) \\
&= E[|e(n)|^2]_{\min} + \mathbf{v}^H(n)\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H\mathbf{v}(n) \\
&= E[|e(n)|^2]_{\min} + \left[\mathbf{Q}^H\mathbf{v}(n)\right]^H\mathbf{\Lambda}\mathbf{Q}^H\mathbf{v}(n) \\
&= E[|e(n)|^2]_{\min} + \mathbf{v}'^H(n)\mathbf{\Lambda}\mathbf{v}'(n)
\end{aligned}
$$

其中， $\mathbf{v}'(n)=\mathbf{Q}^H\mathbf{v}(n)$。

二维情况举例说明几何意义(以实数为例)



- 均方误差超抛物面在水平面上的投影；
- 误差椭圆重心不一定在该坐标的原点；
- v坐标系以误差椭圆的重心为原点；
- v′为主轴：

  平移：$\mathbf{v}(n)=\mathbf{w}(n)-\mathbf{w}_o$；

  旋转：$\mathbf{v}'(n)=\mathbf{Q}^T\mathbf{v}(n)$。

□均方误差指数衰减的时间常数

由 $E\left[\left|e(n)\right|^2\right]=E[|e(n)|^2]_{\min}+\mathbf{v}'^H(n)\boldsymbol{\Lambda}\mathbf{v}'(n)$ 可得

$$E\left[\left|e(n)\right|^2\right]=E[|e(n)|^2]_{\min}+\mathbf{v}'^H(0)\boldsymbol{\Lambda}[\mathbf{I}-\mu\boldsymbol{\Lambda}]^{2n}\mathbf{v}'(0)$$

$$\left(\text{提示：}\mathbf{v}'(n+1)=[\mathbf{I}-\mu\boldsymbol{\Lambda}]\mathbf{v}'(n)\right)$$

结论:

- ✓ 每次迭代，等式右边第二项以$[\mathbf{I}-\mu\boldsymbol{\Lambda}]^2$ 速度衰减;
- ✓ 当 $n\to\infty$时, $E\left[\left|e(n)\right|^2\right]=E[|e(n)|^2]_{\min}$;
- ✓ $E[e^2(n)]$ 随 n增加的衰减比 $\mathbf{w}(n)$ 衰减快一倍。

如果权向量有M个分量，则**权值误差均方误差**的衰减具有M个特征振动模式。

定义$\gamma_m^2=(1-\mu\lambda_m)^2$ 。 如将$\left(\gamma_m^2\right)^n$看成随迭代次数$n$作**指数衰减**，则有

$$\gamma_m^2=[1-\mu\lambda_m]^2=e^{-1/\tau_{mse}}\Rightarrow\gamma_m=1-\mu\lambda_m=e^{-1/(2\tau_{mse})}$$

$$\gamma_m = 1 - \mu\lambda_m = e^{-1/(2\tau_{mse})}$$

采用泰勒展开，上式可表示为

$$\gamma_m = I - \mu\lambda_m = e^{-1/(2\tau_{mse})} = 1 - \frac{1}{2\tau_{mse}} + \frac{1}{2!(2\tau_{mse})^2} + \cdots$$

略去等式右边的高次项，即

$$1 - \mu\lambda_m \approx 1 - \frac{1}{2\tau_{mse}} \quad \Rightarrow \quad \tau_{mse} = \frac{1}{2\mu\lambda_m}$$

- 随 n 增加，$E[e^2(n)]$ 的衰减比w(n)衰减快一倍。

◻决定衰减快慢的主要因素

- ✓ 对应于M个权值的自适应滤波器，其权值误差向量的衰减，是由M个自由衰减模式决定；
- ✓ 在最不利的情况下，具有最大衰减常数的分量对衰减起主导作用；
- ✓ 一般情况下的收敛速度由最大时间常数决定，

- 令 **R** 的最小和最大特征值分别为$\lambda_{\min}$ 和 $\lambda_{\max}$ 。
- 如果按照之前的条件$0<\mu<2/\lambda_{\max}$ 取步长的上界$\mu=2/\lambda_{\max}$，则最大的时间常数为

$$\tau_{\max}=\frac{1}{\mu\lambda_{\min}}\geq\frac{1}{2}\frac{\lambda_{\max}}{\lambda_{\min}}$$

- $\lambda_{\max}/\lambda_{\min}$称为输入信号相关矩阵的特征值扩散度。因此，输入信号相关矩阵的特征值扩散度越大，最大的时间常数 $\tau_{\max}$ 就越大，自适应滤波器收敛就越慢。

- 白噪声对应相关矩阵的特征值扩散度最小，即 $\lambda_{\max}/\lambda_{\min}=1$，因而收敛最快；
- 随着信号的相关性越强，其对应相关矩阵的特征值扩散度越大，因而收敛越慢；

- 因此，LMS算法的主要缺点是，当输入信号的相关矩阵特征值较大时，收敛很慢。
  - **解决思路：将输入信号去相关（白化）**
  - **主要算法：LMS格型算法、RLS算法、仿射投影算法、子带自适应算法等。**

梯度噪声引起的失调

- 理想情况下，在自适应过程中，若没有环境噪声引起的误差，权向量 $\mathbf{w}(n)$ 最终将收敛于均方性能曲面的最小点 $\mathbf{w}_o$；
- 在此条件下，向量 $\mathbf{v}(n)$ 的协方差为0，均方误差 $E[|e(n)|^2]=E[|e(n)|^2]_{\min}=\sigma_v^2$。

- 实际上，一般只能使 $E[\mathbf{w}(n)]=\mathbf{w}_o$，这是因为LMS算法用瞬时梯度代替梯度时会产生梯度估计误差(梯度噪声)。
- 梯度噪声导致 $\mathbf{w}(n)$ 的值随机起伏。这意味着即使滤波器达到了稳态，均方误差 $E[|e(n)|^2]$ 任然大于 $E[|e(n)|^2]_{\min}=\sigma_v^2$ 。

- 定义超量均方误差为 $\zeta = E[|e(n)|^2] - \sigma_v^2$
- 归一化后的超量均方误差，称为失调，即

$$M = \frac{\zeta}{\sigma_v^2}$$

- 该量可以近似为

$$M = \mu \mathrm{tr}[\mathbf{R}]$$

即，失调正比于步长。

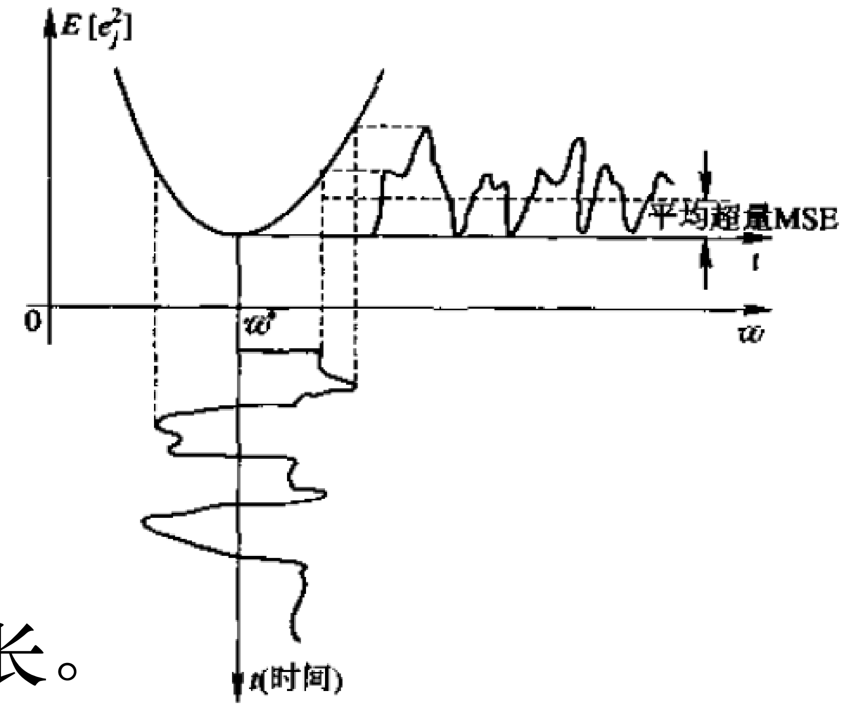


图 4.13　由于梯度噪声而造成的超量均方误差